# BOUNDS ON EXPECTATIONS OF ORDER STATISTICS FOR DEPENDENT SAMPLES

BY

DONALD R. HOOVER

TECHNICAL REPORT NO. 432
JULY 24, 1990

Prepared Under Contract
N00014-89-J-1627 (NR-042-267)
For the Office of Naval Research

Herbert Solomon, Project Director





DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA

## 1. Introduction

Let $X_1, X_2, \ldots, X_n$ be random variables with order statistics denoted by $X_{1:n} \leq X_{2:n} \leq \ldots \leq X_{n:n}$. Bounds on the expected values of order statistics when the $X_i$ are iid with finite mean $\mu$ and finite variance $\sigma^2$ are well known and given in David (1981, Chapter 3). Arnold and Groeneveld (1979) have derived bounds for expectations of order statistics when the identical and independence conditions for the distributions are dropped. Define $\mu_{g:n}$ to be $E[X_{g:n}]$ for $g=1,\ldots,n$ then the Arnold and Groeneveld bound is:

$$(1) \quad \bar{\mu} - \left(\left(\Sigma\, \sigma_i^2 + \Sigma'\mu_i-\bar{\mu})^2\right) \frac{n-g}{n\cdot g}\right)^{\frac{1}{2}} \leq \mu_{g:n} \leq \bar{\mu} + \left(\left(\Sigma\sigma_i^2 + \Sigma(\mu_i-\bar{\mu})^2\right) \frac{g-1}{n(n-g+1)}\right)^{\frac{1}{2}}$$

where the summations are from 1 to n and the bars denote means. When $\mu_i = \mu$ and $\sigma_i^2 = \sigma^2$ for all i then (1) becomes

$$(2) \quad \mu-\sigma \left(\frac{n-g}{g}\right)^{\frac{1}{2}} \leq \mu_{g:n} \leq \mu + \sigma \left(\frac{g-1}{n-g+1}\right)^{\frac{1}{2}}$$

In this paper, new bounds for expected values of order statistics are proposed which do not require finite variances and can take into account known dependence among subsets of k variables for $k < n$.

## 2. The New Bounds

From here on, let $S_j$ be a set of k-1 integers from $1,2,\ldots, j-1$ for $j > k$. Let $j_1 < j_2 < \ldots < j_{k-1}$ be the integers in $S_j$. Now an obvious fact is stated.

FACT 1:

For any $j = k+1,\ldots,n$;

$$\mathrm{Max}(x_j, x_{j_1}, x_{j_2}, \ldots, x_{j_{k-1}}) - \mathrm{Max}(x_{j_1}, x_{j_2}, \ldots, x_{j_{k-1}}) \geq 0$$

This fact is used to prove the following lemma.

LEMMA 1:

$$(3) \quad \mathrm{Max}\,(x_1, x_2, \ldots, x_n) \leq \mathrm{Max}\,(x_1, x_2, \ldots, x_k) +$$

$$\sum_{j=k+1}^{n}\left[\mathrm{Max}(x_j, x_{j_1}, x_{j_2}, \ldots, x_{j_{k-1}}) - \mathrm{Max}\,(x_{j_1}, x_{j_2}, \ldots, x_{j_{k-1}})\right]$$

Proof - Let $n^*$ be the smallest integer such that $x_n^* = \mathrm{Max}\,(x_1, x_2, \ldots, x_n)$

CASE I - $n^* \leq k$

Then $\mathrm{Max}\,(x_1, \ldots, x_n) = x_n^* = \mathrm{Max}\,(x_1, \ldots, x_k)$ and from fact 1, it follows that (3) holds.

CASE II $k < n^* \leq n$

Then $\mathrm{Max}\,(x_1, \ldots, x_n) = x_n^* = \mathrm{Max}(x_1, \ldots, x_k) + \sum_{j=k+1}^{n^*}\left[\mathrm{Max}(x_1, \ldots, x_j)\right.$

$\left. - \mathrm{Max}(x_1, \ldots, x_{j-1})\right]$

$$= \mathrm{Max}\,(x_1, \ldots, x_k) + \sum_{j=k+1}^{n^*}\begin{bmatrix} 0 & \text{if } x_j \leq \mathrm{Max}\,(x_1, \ldots, x_{j-1}) \\ x_j - \mathrm{Max}(x_1, \ldots, x_{j-1}) & \text{otherwise} \end{bmatrix}$$

$$\leq \mathrm{Max}\,(x_1, \ldots, x_k) + \sum_{j=k+1}^{n^*}\begin{bmatrix} 0 & \text{if } x_j \leq \mathrm{Max}\,(x_{j_1}, \ldots, x_{j_{k-1}}) \\ x_j - \mathrm{Max}(x_{j_1}, \ldots, x_{j_{k-1}}) & \text{otherwise} \end{bmatrix}$$

$$= \mathrm{Max}\,(x_1, \ldots, x_k) + \sum_{j=k+1}^{n^*}\left[\mathrm{Max}(x_j, x_{j_1}, \ldots, x_{j_{k-1}}) - \mathrm{Max}(x_{j_1}, \ldots, x_{j_{k-1}})\right]$$

{by FACT 1}

$$\leq \mathrm{Max}\,(x_1, \ldots, x_k) + \sum_{j=k+1}^{n}\left[\mathrm{Max}(x_j, x_{j_1}, \ldots, x_{j_{k-1}}) - \mathrm{Max}(x_{j_1}, \ldots, x_{j_{k-1}})\right]$$

**THEOREM 1:**

An upper bound $\mu_k(\mu_{n:n})$ for $\mu_{n:n}$ is given by:

$$(4) \quad \mu_{n:n} \leq U_k(\mu_{n:n}) = E[\mathrm{Max}(X_1, \ldots, X_k)] + \sum_{j=k+1}^{n}\left[E[\mathrm{Max}(X_j, X_{j_1}, \ldots, X_{j_{k-1}})] - E[\mathrm{Max}(X_{j_1}, \ldots, X_{j_{k-1}})]\right]$$

Provided the above expectations are all defined and finite.

Proof: Taking expectations over (3)

All of the expectations in (4) will be defined and finite if $\mu_i$ is finite for all $i=1,\ldots,n$. These expectations may be defined and finite under other circumstances as well. Note that this bound does not require finite variances.

The concept behind this theorem is similar to that of the Bonferroni inequalities. If each variable $X_i$ takes on only the values 0 and 1 then $X_{n:n}$ can be thought of as $\bigcup_{i=1}^{n} \{E_i\}$ where $E_i$ is the event that $X_i = 1$. When this is the case, the bound given by (4) is the same as the SCAUB improved Bonferroni bound presented in Hoover (1988).

No matter what the ordering of variables $X_1, X_2, \ldots, X_n$ or choice of elements $j_1 < j_2 < \ldots < j_{k-1}$ in $S_j$: $j=k+1,\ldots, n$, (4) is an upper bound. The value of this upper bound does, however, depend on the ordering of variables and choice of elements. In general, the bound given by (4) will be lower if for all $j=k+1,\ldots, n$: (i) The variables represented in $S_j$ have a high positive correlation with variable $X_j$ and/or (ii) The variables represented in $S_j$ are very likely to contain the maximum value from $X_1, X_2, \ldots, X_j$.

Now a lower bound for $\mu_{1:n}$ is given

COROLLARY 1.1:

If all values in (5) are defined and finite, which will be true if $\mu_i$ is finite for all i, then one can produce $L_k(\mu_{1:n})$, a lower bound for $\mu_{1:n}$, of the following form:

$$(5) \quad \mu_{1:n} \geq L_k(\mu_{1:n}) = E[\mathrm{Min}(X_1,\ldots,X_k)] + \sum_{j=k+1}^{n} \Big[E[\mathrm{Min}(X_j, X_{j_1},\ldots,X_{j_{k-1}})] - E[\mathrm{Min}(X_{j_1},\ldots,X_{j_{k-1}})]\Big]$$

Corollary 1.1 immediately follows from applying Theorem 1 to $(-X_1, \ldots, -X_n)$.

The Range of $(X_1, \ldots, X_n)$ is defined as $\mathrm{Max}(X_1, \ldots, X_n) - \mathrm{Min}(X_1, \ldots, X_n) = X_{n:n} - X_{n:1}$. Refer to this as $r_n$. Then an upper bound for $E[r_n]$ is given below.

COROLLARY 1.2:

If all values in (6) are defined and finite, which will be true if $\mu_i$ is finite for all i, then one can produce $U_k(r_n)$, an upper bound for $E[r_n]$, of the following form:

$$(6)\quad E[r_n] \le U_k(r_n) = E[\mathrm{Range}(X_1, \ldots, X_k)] + \sum_{j=k+1}^{n} \Big[ E[\mathrm{Range}(X_j, X_{j_1}, \ldots, X_{j_{k-1}})] - E[\mathrm{Range}(X_{j_1}, \ldots, X_{j_{k-1}})] \Big]$$

Proof:

{By Theorem 1 and Corollary 1.1}

$$E[r_n] = \mu_{n:n} - \mu_{1:n} \quad \le \quad (4) - (5) = (6)$$

It will now be shown that increasing k, the size of the subset, sharpens the bounds given by Theorem 1 and its corollaries 1.1 and 1.2.

THEOREM 2

As before, let $S_j$ be a set of k-1 integers $(j_1 < j_2 < \ldots < j_{k-1})$ from $1,2,\ldots, j-1$ for $j \ge k+1$. Now let $k' = k+1$ and $S_j'$ be a set of k integers from $1,2,\ldots, j-1$ for $j \ge k'+1 = k+2$, with $S_j' \supset S_j$. Let $\tilde{j}$ be the unique integer in $S_j'$ which is not in $S_j$. Finally, define $U_k'(\mu_{n:n})$ to be the upper bound (4) corresponding to $S_j'$, $L_k'(\mu_{1:n})$ to be the lower bound (5) corresponding to $S_j'$ and $U_k'(r_n)$ to be the upper bound (6) corresponding to $S_j'$. As before, assume all expectations are defined and finite. Then:

(i) $\mu_{n:n} \le U_k'(\mu_{n:n}) \le U_k(\mu_{n:n})$

(ii) $\mu_{1:n} \geq L_k'(\mu_{1:n}) \geq L_k(\mu_{1:n})$

(iii) $E[r_n] \leq U_k'(r_n) \leq U_k(r_n)$

Proof of (i)

FACT 2: for all a,b,c

$$[\text{Max}(a,b,c) - \text{Max}(a,b)] - [\text{Max}(b,c) - (b)] =$$

$$\begin{bmatrix} 0 & \text{if } c \leq \text{Max}(a,b) \\ c-\text{Max}(a,b) & \text{if } c > \text{Max}(a,b \end{bmatrix} - \begin{bmatrix} 0 & \text{if } c \leq b \\ c-b & \text{if } c > b \end{bmatrix} =$$

$$\begin{bmatrix} 0 & \text{if } c \leq b \\ 0 & \text{if } c > b \geq a \\ b-a & \text{if } c \geq a > b \\ b-c & \text{if } a > c > b \end{bmatrix} \leq \quad 0 \text{ and thus, whenever it is defined:}$$

$$E[\text{Max}(A,B,C) - \text{Max}(A,B)] - E[\text{Max}(B,C) - (B)] \leq 0$$

but if $S'_{k+1}$ is defined to be $(1,2,\ldots,k)$ and $(\tilde{k+1})$ is defined to be the unique integer from $(1,2,\ldots,k)$ which is not in $S_{k+1}$, then:

$U_k'(\mu_{n:n}) - U_k(\mu_{n:n})$ can be written as

$$\sum_{j=k+1}^{n} \left[ E[\text{Max}(A_j, B_j, C_j) - \text{Max}(A_j, B_j)] - E[\text{Max}(B_j, C_j) - (B_j)] \right]$$

which by Fact 2 is a summation of nonpositive terms and thus must be nonpositive

where $C_j = X_j$

$B_j = \text{Max}\,(X_{j_1}, X_{j_2}, \ldots, X_{j_{k-1}})$ For $j_1, j_2, \ldots, j_{k-1} \in S_j$

$A_j = X_{\tilde{j}}$ where $\tilde{j}$ is the element in $S'_j$ which is not in $S_j$

Proofs of (ii) and (iii) follow from (i) in the same fashion that the proofs of Corollaries 1.1 and 1.2 followed from Theorem 1.

Finally, Theorem 1 and Corollary 1.1 can be extended to produce upper and lower bounds for $E[X_{g:n}]$ where $k < g < n-k+1$ .

**THEOREM 3:**

Upper and lower bounds for $\mu_{g:n}$ are

(7) $\mu_{g:n} \le \mu_{g:g} \le$ Any upper bound for $\mu_{g:g}$ given by (4) under the assumptions of Theorem 1.

(8) $\mu_{g:n} \ge \mu_{1:n-g+1} \ge$ Any lower bound for $\mu_{1:n-g+1}$ given by (5) under the assumptions of Corollary 1.1.

Proof

$x_{g:n} = x_{g:g}$ iff $x_{g:g} \le \mathrm{Min}(x_{g+1},\ldots,x_n)$ . Otherwise $x_{g:g} > x_{g:n}$. Taking expectations yields the inequalities in (7). The proof of (8) is similar.

## 3. Example of Upper Bound for $\mu_{n:n}$

Let $\theta$ be a random angle with $f(\theta) = \frac{1}{2\pi}$ for $-\pi \le \theta < \pi$ and let $X_1 = \mathrm{Cos}(\theta)$, $X_2 = \mathrm{Cos}(\theta+\varepsilon)$, $X_3 = \mathrm{Cos}(\theta+2\varepsilon),\ldots,$ $X_n = \mathrm{Cos}(\theta+(n-1)\varepsilon)$ for $\varepsilon > 0$. Then by (3) with $k=2$ and $j_1 = j-1$ for $j=3,\ldots,n$, it follows that

$$\mu_{n:n} \le U_2(\mu_{n:n}) = E[\mathrm{Max}(X_1,X_2)] + \sum_{j=3}^{n} \Big[E[\mathrm{Max}(X_j,\ X_{j-1}) - E[X_j]\Big]$$

{Since $E[X_j] = 0$ for all j}

$$= \sum_{j=2}^{n} E[\mathrm{Max}(X_j,\ X_{j-1})]$$

{invariance}

$$= (n-1)\ E[\mathrm{Max}(X_1,\ X_2)]$$

(9)
$$= (n-1)\ E[\mathrm{Max}(\mathrm{Cos}(\theta),\ \mathrm{Cos}(\theta+\varepsilon))]$$

{symmetry}

$$= (n-1)\ 2\ E[\mathrm{Cos}(\theta)\ I_{\{\cos(\theta) > \cos(\theta+\varepsilon)\}}]$$

(10)
$$= \frac{4(n-1)}{2\pi}\ \left|\ \sin(\arctan\ (\frac{1-\cos(\varepsilon)}{\sin(\varepsilon)}))\right|$$

For all values of n and $\varepsilon$, it is possible to compute (10) with a hand calculator and easy to obtain. (9) by evaluating a single integral with a computer. At the same time, to calculate the exact value of $\mu_{n:n}$ would require a more difficult integration than needed for (9). Clearly a lower bound for $\mu_{n:n}$ is $E[Max(X_1, X_n)]$. Table 1 gives upper and lower bounds for $\mu_{n:n}$ when $\varepsilon = \pi/180$ for various values of n.

Table 1

Upper and Lower Bounds for the

$$E[Max(Cos(\theta), Cos(\theta + \frac{\pi}{180}),\ldots,Cos(\theta + \frac{(n-1)\pi}{180})] \text{ where}$$

$$f(\theta) = \frac{1}{2\pi} : -\pi \le \theta < \pi$$

| | Lower Bound | Upper Bounds | |
|---|---|---|---|
| (n-1) | $E[Max(cos(\theta), cos(\theta + \frac{(n-1)\pi}{180})]$ | (9) or (10) | Arnold & Groeneveld |
| 5 | 0.027769 | 0.027778 | 1.581139 |
| 10 | 0.055488 | 0.055556 | 2.236068 |
| 50 | 0.269051 | 0.277778 | 5.000000 |
| 90 | 0.450111 | 0.500000 | 6.708204 |

For this problem, Arnold and Groeneveld's distribution free bounds are always larger than 1.0, the maximum value a cosine function can take, and thus unsatisfactory. On the other hand, the new upper bounds proposed in this paper which incorporate known dependencies are quite sharp, even for large n, as is indicated by their closeness to the lower bounds.

## REFERENCES


Arnold, B. and Groeneveld R. (1979), "Bounds on Expectations of Linear Systematic Statistics Based on Dependent Samples," Ann Statist 7, 220-3.

David, H. (1981), Order Statistics, Wiley, New York.

Hoover, D. (1988), "Subset Complement Addition Upper Bounds," submitted for publication.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE *(When Data Entered)*

# REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

| | |
|---|---|
| 1. REPORT NUMBER: 432 | 2. GOVT ACCESSION NO. |
| 3. RECIPIENT'S CATALOG NUMBER | |
| 4. TITLE *(and Subtitle)*: Bounds On Expectations Of Order Statistics For Dependent Samples | 5. TYPE OF REPORT & PERIOD COVERED: TECHNICAL REPORT |
| | 6. PERFORMING ORG. REPORT NUMBER |
| 7. AUTHOR(s): Donald R. Hoover | 8. CONTRACT OR GRANT NUMBER(s): N00014-89-J-1627 |
| 9. PERFORMING ORGANIZATION NAME AND ADDRESS: Department of Statistics, Stanford University, Stanford, CA 94305 | 10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS: NR-042-267 |
| 11. CONTROLLING OFFICE NAME AND ADDRESS: Office of Naval Research, Statistics & Probability Program Code 1111 | 12. REPORT DATE: July 24, 1990 |
| | 13. NUMBER OF PAGES: 10 |
| 14. MONITORING AGENCY NAME & ADDRESS *(if different from Controlling Office)* | 15. SECURITY CLASS. *(of this report)*: UNCLASSIFIED |
| | 15a. DECLASSIFICATION/DOWNGRADING SCHEDULE |

16. DISTRIBUTION STATEMENT *(of this Report)*

APPROVED FOR PUBLIC RELEASE: DISTRIBUTION UNLIMITED

17. DISTRIBUTION STATEMENT *(of the abstract entered in Block 20, if different from Report)*

18. SUPPLEMENTARY NOTES

19. KEY WORDS *(Continue on reverse side if necessary and identify by block number)*

~~Expectation of~~ Order Statistics, ~~Upper and Lower Bounds~~, statistical ~~Dependent~~ Samples, ~~Known~~ Dependencies, ~~Expected Range.~~ Mathematical Functions.

20. ABSTRACT *(Continue on reverse side if necessary and identify by block number)*

PLEASE SEE FOLLOWING PAGE.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

S/N 0102-014-6601

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE *(When Data Entered)*

TECHNICAL REPORT NO. 432

## 20. ABSTRACT

Arnold and Groeneveld (1979) presented distribution free bounds for $E(X_{g:n})$ the expected value of the 'gth' order statistic when sampling 'n' possibly dependent observations from populations each with finite expectations mu(i) $\mu_i$ and finite variances sigma squared (i) $\sigma_i^2$. Here, the problem of incorporating known dependencies among the distributions of observations in a sample to produce bounds for the order statistics of that sample is considered. Limits are obtained for expectations of the 'gth' order statistic and the range. These limits hold even when the population variances are not finite. An application of the upper bound for $E(X_{n:n})$ is given. Keywords; Block 19, 1473, page